

Weekly report

1 Done

1.1 RelationLine 软文（modification）

听说你也想化身福尔摩斯？“关系线”了解一下

从一份数据中找到蛛丝马迹已经不是一件难事。然而，在解决真实问题的时候，可以利用的数据往往不止一份。当线索一条条摆在面前，要如何从中整理出唯一的真相呢？近日，阿里巴巴-浙江大学前沿技术联合研究中心发表了最新研究成果关系线（RelationLine）——一个可以从异构城市数据中推理出个体关系网络的可视分析系统。

不同的城市数据可以为个人关系网络的构建提供不同的线索。比如，功能区的分布信息可以用来推测人们前往特定地理位置的目的；社交媒体数据可以提供已发生事件的文字描述；轨迹数据可以揭示不同的伴随关系。伴随关系是指两个以上的个体在一段时间内停留在同一地点或有相同的行进轨迹。这种关系在不同个体间、不同时间、不同地点，都会对应不同的解释。例如，人和车的短时间伴随关系可能意味着这个人是乘客，长时间伴随关系则意味着这个人是司机。

从轨迹数据中挖掘伴随关系有着很大的应用价值，比如向上班族推荐可以一起拼车去公司的人；构建个人关系网络等等。然而，这也很具有挑战。轨迹数据由于数据来源不同，精度不同，存在着难以匹配的问题。此外，伴随关系可能具有多种意义，这就导致解释它们背后的含义十分困难。

针对这两大挑战，浙江大学计算机辅助设计与图形学国家重点实验室的陈为老师团队和阿里云城市大脑团队合作完成了一个可视分析系统——关系线。针对第一个挑战，关系线给出了一种基于四叉树的轨迹匹配方法。该方法首先将地图上的点通过四叉树编码成一长串数字。在匹配的过程中，用户可以通过设定匹配位数来调整匹配的精确度，以适应不同数据、分析的匹配需求。接下来，根据数据的不同，关系线提供两种计算轨迹之间相似度的方法。对于签到、打卡等只记录了几个轨迹点的数据，关系线统计可以匹配到的轨迹点个数；对于比较完整的轨迹数据，关系线计算每段轨迹 r 的匹配时长（如图 1）总和（见下公式）。这样一来，不同精度、格式的数据都可以进行相互匹配。

$$DoM(r_i, r_j) = \begin{cases} 0, & \text{地点不一致} \\ 0, & \text{时间不一致} \\ 1, & \text{数据只包含轨迹点} \\ T_{\text{重合部分}}, & \text{其他} \end{cases}$$

$$DoM(R1, R2) = \sum_{r_i \in R1 \text{ 且 } r_j \in R2} DoM(r_i, r_j)$$

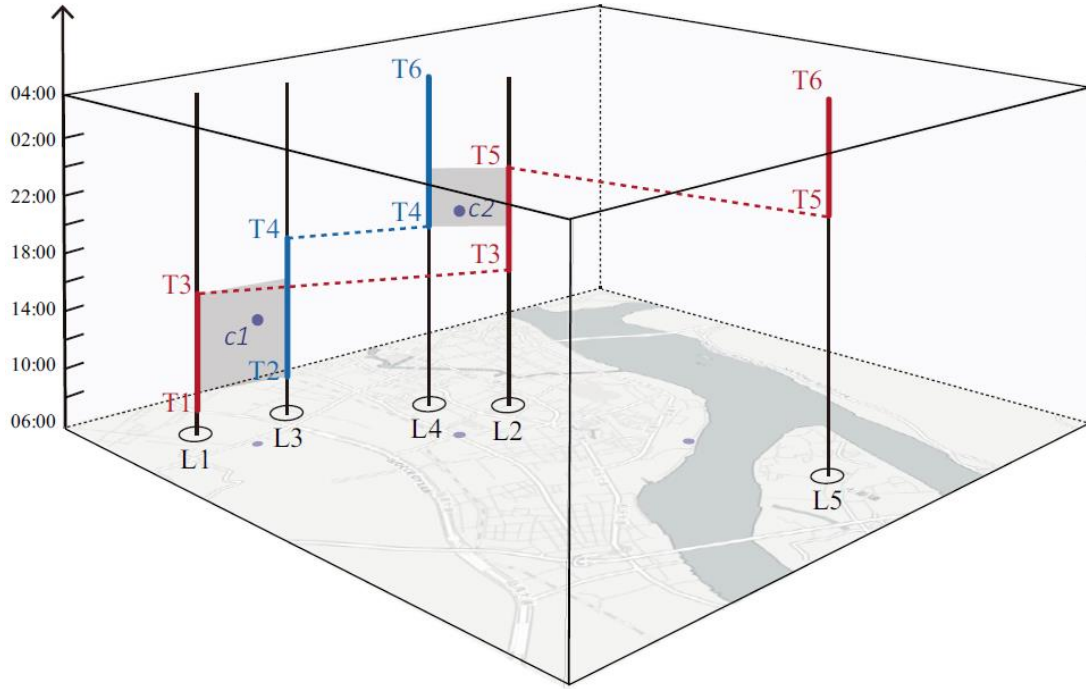


图 1 轨迹的时空匹配方法

针对的第二个挑战，该研究给出了一套具体的分析方法。下面，我们将通过一个关于出租车轨迹的例子来介绍这个方法。首先，用户需要从数据库中选择一个分析目标，这个目标可以是人的轨迹，也可以是车的轨迹。在这个案例中，我们选择了一辆几乎昼夜不间断工作的出租车（图 2（1）中的绿色块代表在相应时间段内该出租车上乘客）。我们可以推断，这辆出租车有不止一位司机，他们会交替驾驶出租车。

接下来，为了找到他们的交班规律，我们需要通过匹配出租车和手机基站记录的轨迹来找到相应的司机。选择合适的轨迹段匹配可以提高效率，也可以避免其他路人的干扰。观察绿色块的分布，我们发现在每天早上 7 点左右都会有一段不接单的时间。可以推测这段时间是司机的交班时间之一。因此，我们分别选择在这段时间的左侧（夜班司机）和右侧（白班司机）附近的轨迹段进行匹配。其中，综合匹配度最高的是一条和出租车轨迹一直处于吻合状态的轨迹

（在图 2（1）中，匹配部分用灰色线段标于绿色块上方）。可以推断它是司机师傅们用来导航的固定车载设备。接下来的两位，则是我们要找的两司机了（在图 2（1）中，搜索到的两位司机的匹配部分，分别用蓝色和橙色线段标出）。但是，在确认完整轨迹匹配程度的时候，我们注意到，白班的这位司机轨迹只能匹配到出租车前六天的轨迹。因此，肯定有第三位司机在另外三天驾驶了出租车。继续对这部分轨迹进行匹配（如图 2（2）），我们最终找到了第三位司机（在图 2（1）中，匹配部分用紫色线段标出）。

最后一步是对三位出租车司机的交接班行为进行分析。当轨迹匹配部分都被标出后，我们可以看出两个交接班时间分别是早上 7 点左右和晚上 17 点半左右。在地图上观察具体轨迹，我们定位到了这两个交接班时间对应的交班地点（如图 3 和图 4）。

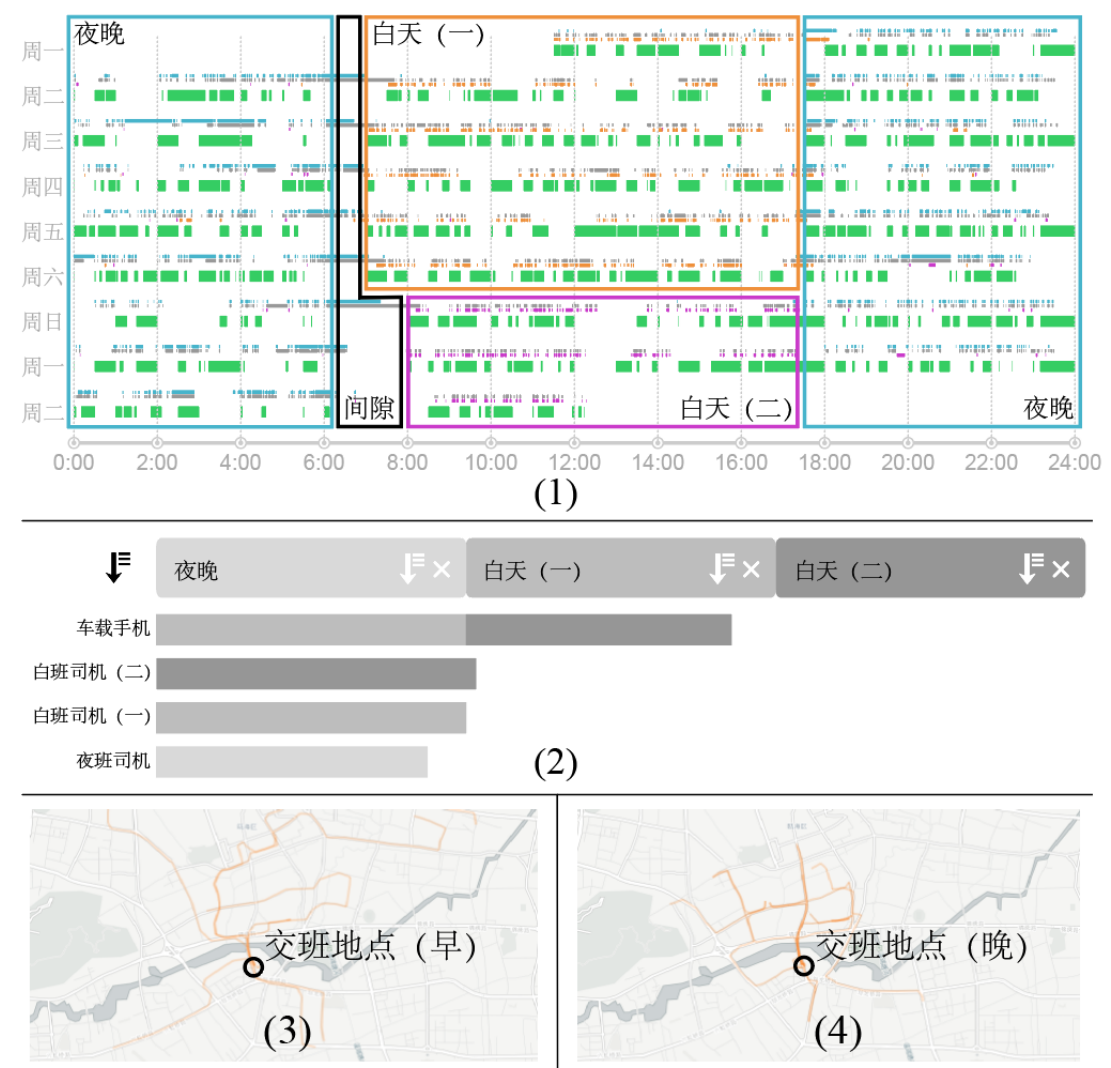


图 2 出租车案例分析

注 1：论文中数据为实验室数据，非阿里内部或城市大脑的数据。数据均经过脱敏处理。
论文 RelationLines: Visual Reasoning of Egocentric Relations from Heterogeneous Urban Data，
将发表于 ACM Transactions on Intelligent Systems and Technology，2018，主要作者包括陈
为、夏菁和王叙萌等。

1.2 Monography Modification

（原有内容 修改内容 增加内容）

14 可视化评测

修改全文语句不通顺、介绍不清楚的地方

14.1 评测的流程

14.2 影响评测效度的因素

14.2.1 参与用户

被试应尽可能覆盖可能的目标用户范围，分布尽可能均匀

14.2.2 测试任务

任务难度设定

14.2.3 数据类型

14.2.4 实验环境（时间控制、设备、系统设计等）

14.2.5 评测的指标

14.3 评测的方法

14.4 评测实例分析

压缩原有例子

增加例子[2017 TVCG] An Evaluation of Visual Search Support in Maps （涉及眼动仪的使用，p-value 指标估计等新内容）

14.5 总结

14.6 参考文献

1.3 Paper reading

[2017 EuroVis] **Graph layouts by t-SNE**

[2017 InfoVis] **Nonlinear Dot Plots**

2 Progress

Item	Deadline	Current progress	Remark
Patent	4.27	-	Me: a patent for the approach; Rusheng: a patent for the privacy-preserving model.
Monography	5.10	Plan already	-

		given.	
Privacy program	10.31	Surveying.	Graph privacy preservation though differential privacy.